Problem-1

To solve this problem I decided to write a code in python language called “getBlogs.py” to get the 100 unique blogs.

Libraries used:

import requests

I wrote a “getUniqueBlogs” function to retrieve 100 unique blogs. Inside that function I am creating a “100BlogUrls.txt” text file to store 100 blogs. To collect unique element in python we have one data type called “set” and I used sets to collect my 100 unique blogs. For getting blog I am hitting this URL <http://www.blogger.com/next-blog?navBar=true&blogID=3471633091411211117> using “requests” library and this is looping 100 time through while loop and storing in “uniqueBlogs” variable. And after collecting 100 blogs, I added two URL given in the question to “uniqueBlogs” variable. The reason for collecting 100 blogs instead of 98 is sometimes few blogs will be having with no type called “application/atom+xml”. For safety reasons I am collecting 100 blogs.

After collecting 100 unique blogs in “uniqueBlogs” variable. I am getting raw HTML for each and every blog in **uniqueBlogs** . All raw HTML are storing in “blogs” folder. I saved blog id and the URI found inside a file called “**100BlogUrls.txt**”.

Run on the command line

Python getBlogs.py